2013

Using Math Curriculum Based Assessments to Predict Student Performance on the Pennsylvania System of School Assessment Math Test

Joseph H. Lucas
Philadelphia College of Osteopathic Medicine, jklucas157@verizon.net

Philadelphia College of Osteopathic Medicine

Department of Psychology

USING MATH CURRICULUM BASED ASSESSMENTS TO PREDICT STUDENT PERFORMANCE ON THE PENNSYLVANIA SYSTEM OF SCHOOL ASSESSMENT MATH TEST

By Joseph H. Lucas

Submitted in Partial Fulfillment of the Requirements for the
Degree of Doctor of Psychology

February 2013

PHILADELPHIA COLLEGE OF OSTEOPATHIC MEDICINE
DEPARTMENT OF PSYCHOLOGY

Dissertation Approval

This is to certify that the thesis presented to us by Joseph Lucas on the 13th day of July, 2012, in partial fulfillment of the requirements for the degree of Doctor of Psychology, has been examined and is acceptable in both scholarship and literary quality.

Committee Members' Signatures:

George McCloskey, PhD, Chairperson

Yuma Tomes, PhD

Dennis L. Seaman, EdD

Robert A DiTomasso, PhD, ABPP, Chair, Department of Psychology

Acknowledgements

Over the course of my academic career and personal life there have been many obstacles; none greater than the completion of my dissertation. In the end, I persevered and achieved my life-long dream. This was a personal and intellectual journey that challenged me to the greatest extent and without the presence of extraordinary individuals in my life, none of this would have been possible.

First, I must thank from the bottom of my heart, the late Joseph McManis. You were not only my first mentor and grandfather, you were my friend. Saying good bye was difficult; I will never forget you.

To my mother, who set an example of excellence, determination and hard work, you supported me and believed in me from the outset. I always want to make you proud. My father, who never accepted anything less than my best, insisted that I succeed, and humbly took his lunch box into that steel mill daily, in the hopes that his son could do better and work in a safe and healthy environment. My two sisters, you were like mothers to me and spent countless hours teaching, caring for and befriending me; you were always my allies.

Kristy, my wife, mother to my three daughters, best friend and lifelong love, you never flinched in the face of adversity and supported my personal and academic endeavors whole-heartedly. We grew up together, made a life, a family and a future; I would not change a minute of the last eighteen years. To my three beautiful daughters, Taylor, Lauren, and Katie, the greatest miracles of my life, I never new beauty or love

until I met you. I am grateful to god every day that I have the privilege of teaching and loving you. Each day and night that I was not home due to my academic and professional endeavors was difficult for me and I missed you dearly. Please accept my apology and learn from my experience; hard work does pay off. Never give up; ever!

To Dr. George McCloskey, the most intellectually gifted and wise teacher of my lifetime. You supported me, kept me on track and challenged me. You will never know the debt of gratitude I owe you. To Dr. Yuma Tomes and Dr. Dennis Seaman, my other committee members, your ideas, feedback, support and participation throughout this process were vital and appreciated. Dr. Rosemary Menutti you supported me tremendously and always lent a caring ear. You are one of the most kind individuals that I know.

To the other colleagues, friends and family members that I did not mention previously, you are not forgotten. Thank you for everything, especially the encouragement, support, laughs and good times.

Abstract

Today, a considerable emphasis is placed on students' performance on state-wide achievement tests. In light of the mounting pressure for accountability for student academic achievement on state-wide tests, the use of curriculum-based measurement (CBM) methods for monitoring student progress, identifying students at risk for failing state tests, and identifying skill deficits to be addressed through interventions to increase student performance could prove beneficial. This study examined the relationship of the Measures of Academic Progress Math CBM (MAP) to performance on the Pennsylvania System of School Achievement (PSSA) with correlational analysis and the calculation of sensitivity and specificity indices and kappa values to examine the predictive power of MAP scores. The study also introduced the use of two additional indices – The Improvement Index and the Instability Index –to describe the relationship between progress monitoring measures (MAP) and outcomes measures (PSSA) and evaluate the effectiveness of instruction and progress monitoring efforts. The study also conducted a more in-depth analysis of score change patterns, analyzing the patterns produced by students' individual score changes from fall MAP to Spring MAP to PSSA.

Table of Contents

List of Tables

Chapter 1

Introduction

Today, a considerable amount of emphasis and importance are placed on a student's performance on state-wide achievement tests. In light of the mounting pressure and accountability for student academic achievement on state-wide tests, the use of curriculum-based measurement (CBM) methods for monitoring student progress, identifying students at risk for failing state tests, and providing appropriate interventions to increase student performance could prove beneficial (McGlinchey & Hixson, 2004). A form of academic achievement assessment, CBM can assist in identifying students at risk for failure by comparing their performance on such measures to a standard of performance. According to Deno and Fuchs, CBMs utilize standardized methods to capture a student's present level of mastery of curricular materials and to compare the student's performance to some predetermined criterion (as cited in Sattler, 2001).

Statement of Problem

An expansive body of literature exists for the support of CBM as an effective tool in predicting success on state-wide reading achievement tests (Barger, 2003; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006; Shaw & Shaw, 2002; Stage & Jacobsen, 2001; Wood, 2006). A considerable amount of research has been devoted to reading achievement in general, particularly in regard to the use of CBM and prediction of achievement on state-wide tests. However, a student's performance on state-wide math assessments is just as, if not more, important.

Recently, school districts effectively implemented a form of CBM, called research-based measurement (RBM), which utilizes benchmark assessment to allow classroom teachers to identify and intervene for students at-risk for failure on state-wide tests.

Determining if a student is at-risk for failing the state-wide math achievement test could be useful in providing early intervention and influencing educational programming (Shapiro et al.; Hintze & Silberglitt; McGlinchey & Hixson; Stage & Jacobsen). However, results need to be replicated for further validation and generalizability. This validation and generalizability should include different examiners, Math CBM measures, state tests, populations, and numbers of students (Hintze & Silberglitt). The current study examines the overall effectiveness of progress monitoring utilizing Math RBM in predicting performance on state-wide math tests. The use of this different type of CBM measure as a predictor enables this study to keep up with current assessment trends and allows for a variety of measurements not previously studied vigorously. Additionally, this study seeks to address which student RBM can and cannot successfully predict students who are and are not at-risk for failure on state-wide achievement tests.

Purpose of the Study

The current study seeks to understand the relationship of Math CBMs to performance on state-wide math assessment. This relationship includes gaining a better understanding of the variability between different Math CBMs and performance on state-wide math assessments, such as the strength of the relationship. Different CBM measures as predictors, particularly RBMs were examined. Factors that indicate why a

Math CBM is a good predictor of performance for some students and not for others were also examined.

Chapter 2

Literature Review

Curriculum-Based Measurement

CBM may be a useful tool for school districts that want to increase state-wide achievement testing scores, as it can provide useful information for instructional decision-making, including a student's readiness to go to the next level of instruction, the skill deficits that need to be addressed, and the tools that might best help the student to obtain or increase the desired skills (Sattler, 2001). Deno and Fuchs pointed out that CBM has been shown to have the ability to improve the match between testing and teaching, assess the performance of a student within the curriculum, determine the effectiveness of current instructional methods, and improve communication between regular and special-education teachers and between teachers and parents (as cited in Sattler, 2001). Additional CBMs performed by school professionals can be used to conduct motivation assessments, error analysis, and progress monitoring to evaluate a student's response to intervention (Ardoin et al., 2004). CBM has been shown to be useful throughout the grades, not just early on in a child's education (Hosp & Fuchs, 2005). With research that supports CBM as a tool for school psychologists and other professionals to use throughout the grades, CBM can be used to screen, diagnose, and monitor a student's academic achievement (Hosp & Fuchs).

Deno and Mirkin (1977) are credited with originating CBM as a method to monitor a student's academic progress. CBM was developed as a measurement system to test the effectiveness of a special-education intervention model (i.e., data-based program modification) by obtaining valid and reliable repeated measures of a student's academic performance in order to evaluate and improve instruction (Deno, 1985; Deno, Fuchs, Marston, & Shin, 2001). In educational decision-making, CBM is used for screening, identifying, and referring students at risk for academic failure; gauging a student's responsiveness to interventions; evaluating the effects of interventions; making instructional decisions; and, most recently, predicting a student's achievement on high-stakes assessments (Deno, 2003; Fuchs, 2004; Fuchs & Deno, 1992; Fuchs, Fuchs, & Courey, 2005).

With the reauthorization of the Individuals with Disabilities Education Improvement Act (IDEIA) in 2004, Response to Intervention was included as an alternative method of identifying students with learning disabilities. Response to Intervention is predicated on the idea that all children receive high-quality instruction in general education classrooms. Children who do not make adequate progress despite high quality instruction as determined by ongoing assessment then are provided with increasingly intense, multitiered interventions that may eventually result in special-education placement. Consequently, Response to Intervention models require progress monitoring or frequent assessment of student performance to make appropriate instructional decisions for children (Fuchs & Fuchs, 2006; Speece, Case, & Molloy, 2003; Wallace, Espin, McMaster, Deno, & Foegen, 2007).

CBM was developed by Stan Deno at the University of Minnesota's Institute for Research on Learning Disabilities during the mid 1970s. Originally developed as a metric to examine the rates of growth in students participating in special education (Stecker, Fuchs, & Fuchs, 2005), its current applications have broadened to both formative and summative assessments of student skill acquisition. Research by Shinn (1989) provided solid validation for CBMs role in monitoring student progress and making subsequent educational decisions about instructional content and strategies (Cusumano, 2007). As can be seen, CBMs qualities also align with accountability requirements as enmeshed in No Child Left Behind legislation (NCLB, 2001). Notably, Fuchs (2004) have addressed this accountability issue and paved a pathway along which CBM data could serve as indices documenting Annual Yearly Progress for students across kindergarten through sixth grade. Application of CBM as an index of Annual Yearly Progress is forthcoming (Cusumano, 2007).

One popular model of Math CBM taps students' fluency with basic math skills. During the administration of Math CBM, students are presented with probes containing either single or mixed-skill calculation problems (Clarke & Shinn, 2004). Directions inform students to complete as many of the problems as possible within the prescribed time limit. Time limits vary based on grade level but vary between 2 to 4 min, with younger students in first through third grades provided with 2 min, to complete math probes presented (Clarke & Shinn, 2004). The number of digits correct in the student's answers becomes the Math CBM score. Students completing the Math CBM are directed to try each item; however, they are allowed to draw an "X" through any problems that they cannot complete. As with other CBM indices, strong documentation of the internal consistency (Fuchs, Fuchs, &Hamlett, 1994), test-retest, and interscoring agreement of Math CBM have been offered (Cusumano, 2007).

As can be noted, CBM has gained strong footing as a metric for monitoring student academic progress, particularly during the elementary years of schooling. Thus, all individuals who work with children must understand its use in the school setting both as a screening instrument that identifies risk levels associated with individual students and as a metric that monitors a student's acquisition of skills. Specifically, this discussion was aimed at increasing practitioners' awareness and understanding of progress monitoring and screening tools upon which student achievement can be assessed. It is hoped that this discussion has increased the knowledge of all who work with children, thus opening channels of communication so that precious energies can be directed toward identifying the method that works best for *all* children (Cusumano, 2007).

Shifting of Current Trends in Progress Monitoring

As demonstrated, CBM is an effective way to monitor a student's progress over the course of a school year. It also has been effective in assisting in differentiating instruction and has been utilized as a predictor of performance on state-wide achievement tests. However, today many school districts across the country are utilizing RBMs, which utilize benchmark assessments. Educators view benchmark measures reflective of such external tests as potentially more valid in making differentiated-instruction decisions that can lead to gains in student learning, higher scores on state standardized tests, and improvements in school-wide achievement (Baenen et al., 2006; Baker & Linn, 2003).

Math as a Critical Skill in the Adult Workforce

With the huge growth of jobs in the technology sector since 1990, math has become a critical skill for Americans seeking employment. Proficiency in mathematics is a prerequisite skill for individuals seeking employment in the fertile technology sector (U.S. Department of Labor, Bureau of Labor Statistics, 1997). National studies indicate that American math students will not have the necessary skills to meet the challenging and changing demands of the United States workplace (Reese, Miller, Mazzeo, & Dossey, 1997). In addition, concerns have been raised by educators, particularly in official U.S. Department of Education (1998) reports. Hiring practices by companies also may be of interest to current math students and educators, as individuals who are proficient in math earn 38% more than individuals who are not.

Although math has always been a core subject in the school curriculum, learning outcomes in math are still not optimal. In the late 1980s, the United States ranked at the bottom in international comparisons among developed countries (Fuchs et al., 1994). The performance of American eighth-grade students in the International Evaluation of Educational Achievement was more than 2 years behind that of such students in high-scoring countries (Fuchs & Fuchs, 2001). The report of the Trends in International Mathematics and Science Study in 2003 showed that the mathematics performance of American eighth-grade students fell behind that of 15 of the 46 participating countries (National Center for Education Statistics, 2003). Recent studies have revealed that the current math performance level of American students has not met the challenging needs of the job market at this time (Clark & Shinn, 2004). Therefore, student performance in mathematics has caused concern among educators in the United States.

Types and Predictive Power of Math CBM

Several Math CBM studies have been focused on the measurement of early math skill acquisition of students in kindergarten and first grade. These studies utilized measures based on the academic and cognitive-processing difficulties presented in the literature (Fuchs et al 2007). These studies utilized single-skill probes (probes that measure one skill at time, such as addition or subtraction) or multiple-skill probes (measuring two or more math skills, such as addition, subtraction, multiplication, and division in one probe), with the latter providing much better predictive power. These studies indicated modest predictive power, as a set of single-skill probes and multiple-skill probes was utilized to predict performance on standardized tests of academic achievement. Those studies that utilized specificity and sensitivity showed greater specificity than sensitivity. That is, Math CBM measures showed a greater overlap when math achievement was in the Proficient range and the Math CBM measures indicated No-Risk than when math achievement was in the Not Proficient range and the Math CBM measure indicated At Risk. The Math CBM measures in the studies were better at identifying students who earn Proficient scores on math achievement tests than at identifying students who would not earn Proficient scores on math achievement tests.

Four observations or shortcomings were noted by Fuchs et al. (2007) on the summative work of studies utilizing CBM in early elementary grades (K and 1). First, the more complex the screener or set of probes, the stronger the correlations were with outcomes. Not surprisingly, more complex screeners are more likely to predict math achievement and appear to be a sound concept, as math is a complex academic subject with a variety of prerequisite skills and foundational knowledge. Second, the studies of math in early elementary grades may be less likely to predict students who will develop math difficulties later in school. This prediction may be expected, as no cognitive-processing measures were utilized in the studies measuring specificity and sensitivity. Also, various cognitive abilities and math skills (computation vs. story problems) are required to perform the various math problems included on math achievement tests. Third, these screeners may not have sufficient sensitivity to determine fine discrimination of students who have significantly deficient skills as a result of not being sufficiently difficult. Finally, of the existing studies, only one study examined outcomes of 1 year and none analyzed outcomes greater than one year or beyond first grade.

Fuchs et al. (2007) attempted to address aforementioned shortcomings of the investigations. These authors investigated the ability of Math CBM to predict math disabilities and predictive and discriminate validity of monitoring math progress from the beginning of first grade to the end of second grade. Math disabilities were categorized into two groups: math disability-calculation and math disability- word problems, even though this categorization does not appear to encapsulate the math disability subtypes reviewed earlier. The Math CBM in this study was comprised of four tasks: number identification/counting, fact retrieval (math fluency), a multiple-skills math concepts and applications screener, and multiple-skill concepts and applications probe. The Math CBM provided a good fit in helping to identify math disability-calculation and math disability-word problem, with the computation and the concepts and application providing the best accuracy for predicting the math disabilities. CBM computation demonstrated the strongest validity for progress monitoring. These authors discussed at length the practical utility of a multiskill Math CBM as opposed to a single-skill Math CBM because of the wide array of math skills taught in a given curriculum (Fuchs et al., 2007). Christ and Vining (2006) agreed that more complex measures equate with better predictability and achievement trends.

Foegen, Jiban, and Deno (2007) at the Research Institute on Progress Monitoring found only 32 studies that addressed math CBM. None of the studies addressed the problem of establishing parallel forms that would ensure equivalence of scores across multiple assessments. Technical adequacy usually was indicated by reporting reliability and validity data, including internal consistency, test-retest reliability, alternate form reliability, and concurrent and predictive criterion validity.

A recent study conducted by Shapiro et al. (2006) utilized Math CBMs to predict student performance on the Pennsylvania high stakes, standardized achievement test, the Pennsylvania System of School Assessment (PSSA; Pennsylvania Department of Education [PDE], 2002). The math computation and math applications Math CBMs utilized in this study demonstrated strong correlations with midyear assessments as well as with published norm-referenced achievement tests in two districts in Pennsylvania.

The Shapiro et al. (2006) study is the only known study to date to correlate Math CBM performance with the PSSA. This study provides evidence of the practical utility of CBM for determining which students are at risk for poor performance on a state-wide standardized test of achievement, as well as for helping to provide academic intervention to a specific group of students who are experiencing math difficulties in order to improve future performance on PSSAs.

In recent years, school districts across the country have moved to a computerized CBM, called RBM, which saves time in administration and data collection. A commonly used RBM is the Measures of Academic Progress (MAP, NWEA, 2003). In addition, this type of CBM was developed to overcome the limitations of existing Math CBMs and requires validation as a tool. This study seeks to understand students' performance on the Math MAP and the way educators can utilize student performance on the Math MAP to predict who is and is not at risk for failure on the PSSA.

Multicultural Issues Surrounding Math CBM

The United States is currently the most linguistically, culturally, religiously, and ethnically diverse nation in world history (Prewitt, 2002). According to the U.S. Department of Education, the 5.4 million students who have limited English language proficiency represent the fastest growing student population, expected to make up one in every four students by 2025 (ED. Gov, 2006) Therefore, understanding the ways multicultural issues affect results of Math CBM and Math CBM's capability to predict success on high-stakes tests is essential. Although the available research supports Math CBM as a predictor of success on state-wide math achievement tests (Shapiro et al. 2006), it does not address the issue of using Math CBM as a predictor of success for English language learners. However, the use of CBM for English language learners is research supported by the work of Graves, Plascencia-Peinado, Deno and Johnson (2005).

Curriculum based assessment has significant multicultural advantages, as its goal is to measure a student's progress in a curriculum, regardless of age, gender, ethnicity, or religious affiliation. In a series of workshops entitled "The Use of IQ Tests in Special Education Decision Making and Planning" and commissioned by the National Research Council (Morison, White, & Feuer, 1996), the consensus of the expert panelists was that IQ tests continue to play a disproportionate role in eligibility. Thus, assessment tools that inform educators about the nature and effectiveness of instructional interventions need to be part of the special-education eligibility process and to monitor their progress. As conceptualized by Messick (1984), disproportionate representation of an ethnic group is problematic when students are unduly exposed to classification because they receive a poor-quality regular education that hinders educational progress. Accordingly, eligibility assessment methods must systematically eliminate problematic sources of unfair overrepresentation by informing and measuring the effectiveness of educational programs for individual students. A critical aspect of eligibility assessments is the determination of need for services. Therefore, assessment methods are tools that are sensitive to the instructional enterprise and curricula, and are needed to supplement traditional methods of assessment, which primarily have addressed the classification aspect of eligibility decisions (Elliot & Fuchs, 1997).

Although there are several advantages of CBA, there may be a misconceptualization of the difficulties of over identification. Overidentification is likely to be more of a symptom of overreliance on ability-achievement discrepancy and of poor clinical interpretation of intelligence test data. There is an overreliance on the use of full-scale IQ in determining ability-achievement discrepancies and understanding of patterns of subtest performance of individuals from various cultures. Also, CBA and CBM have not been used historically to document that a student’s academic deficits are not the result of a lack of instruction. Therefore, utilizing CBA during the prereferral process to document or rule out a lack of appropriate instruction and sound clinical interpretation of cognitive assessments is most likely to identify appropriately students for special-education programs (Hale, Kaufman, Naglieri, & Kavale, 2004).

Although eligibility decisions are vital aspects of the educational process, this study is concerned with identifying students who are at risk for failure on state-wide achievement tests in the area of mathematics. Multicultural issues that surround mathematics are less predominant in mathematics, as computation and mathematics fluency have very low cultural loads. Mathematical applications and story problems, which have a heavy load on linguistics, present much greater challenges with regards to English language learners. Therefore, Math CBMs that assess math computation and fluency will be highly relevant and will predict student performance on tasks measuring math computation on state-wide achievement tests. With regards to math application and story problems, a student's language proficiency will impact his or her ability to be successful on Math CBMs, as well as on state-wide math achievement tests that assess reasoning and have a heavy language load.

Concluding Summary of the Literature

Currently, the pressures for children to be successful on state-wide achievement tests are considerable and mounting. To this date, substantial amounts of time and money have been spent on research demonstrating the efficacy of reading CBM. A large body of expanding research links CBM reading measures to student performance on state-wide reading achievement tests (Barger, 2003; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Shaw & Shaw, 2002; Stage & Jacobsen, 2001). Although not researched as heavily, but of equal or greater importance, is linking Math CBM measures to student performance on state-wide mathematics achievement tests. As the importance of mathematic skills grows with our advancing technological society, mathematics achievement is becoming a topic of primary focus for educators and researchers.

CBM is a widely accepted and relatively accurate form of assessment that can be utilized to identify students who are at risk for academic difficulties, to monitor students' academic progress, to identify students at risk of failing state tests, and to isolate skill deficits, thereby enabling educators to provide appropriate interventions to increase student performance (McGlinchey & Hixson, 2004). CBMs utilize standardized methods to capture students' present levels of mastery of curricular materials and to compare student's performance to some predetermined criterion (Deno & Fuchs, 1987).

A number of studies have offered some degree of support for the use of Math CBM to identify students who are at risk for academic failure (Christ & Vining, 2006; Foegen, Jiban, and Deno, 2007; Fuchs et al., 2007) and failure on state-wide achievement tests (Shapiro et al., 2006). The body of research appears to support hypotheses that Math CBM can, in fact, be used as a reliable and valid method of predicting which students are at risk for failing state-mandated mathematics tests. The literature to date, however, has not addressed the question of whether or not other forms of assessment might be more effective as predictors of state-wide achievement performance. Although Math CBM measures generally require minimal time to administer, perhaps more in-depth math assessments would provide even better information regarding future state-wide math test performance. Therefore, the investigation of more in-depth assessments for predicting high-stakes, state-wide math testing performance is warranted.

The RBM, a type of CBM, has been utilized widely across the country to assess and monitor student progress. Owing to high reliability and validity and convenience of large group assessment and data collection and analysis, the Math MAP has been utilized readily (NWEA 2007). The Math MAP is a type of RBM that is commonly utilized by school districts across the United States. The goal of this study is to analyze student performance across grade levels and determine if the math MAP is a successful predictor in determining student that are or are not at risk for failure on the PSSA.

Also, there appears to be a need for research that examines the variability between math CBM and student performance on state-wide reading achievement tests. It is important to understand the effectiveness of CBM at predicting performance and the reasons such measures predict performance for some students and not others. Although research indicates that Math CBMs are somewhat effective tools for identifying students who are at risk of failure, the use of Math CBM to predict performance on state-wide math assessments for both high and low socioeconomic status students, as well as English language learners, appears to be an area in need of further research.

Research Questions

Research Question 1: By grade level within each cohort, what is the relationship among scores on the fall and spring administrations of the MAP for math and the PSSA math assessment based on correlation coefficients?

Research Question 2: By grade level within each cohort, what proportion of students was identified as Proficient on the spring administration of the PSSA math assessment?

Research Question 3: By grade level within each cohort, what proportion of students was identified as Not Proficient on the Fall or Spring MAP math assessment and judged as At-Risk of being Not Proficient on the PSSA math assessment?

Research Question 4: By grade level within each cohort, what is the relationship between the Fall and Spring MAP for math proficiency decisions and PSSA proficiency decisions based on multiple indices?

a. At each grade within each cohort, of the students who earned Not Proficient level scores on the PSSA math assessment, what percentage of these students also earned Fall or Spring Math MAP scores in the Not Proficient range (operationally defined as the Sensitivity Index)?

b. At each grade within each cohort, of the students who earned Proficient level scores on the PSSA math assessment, what percentage of these students also earned Fall or Spring Math MAP scores in the Proficient range (operationally defined as the Specificity Index)?

c. At each grade within each cohort, what is the percentage of improvement over chance represented by the relationship between PSSA math assessment level score categories and Fall and Spring Math MAP score level categories (operationally defined as the Kappa Index)?

d. At each grade within each cohort, of the students who earned Not Proficient level scores on the Fall and Spring MAP for math assessment, what percentage of these students also earned Spring PSSA math assessment scores in the Proficient range (operationally defined as the Improvement Index)?

e. At each grade within each cohort, of the students who earned Proficient level scores on the Fall and Spring MAP for math assessment, what percentage of these students also earned Spring PSSA math assessment scores in the Not Proficient range (operationally defined as the Instability Index)?

Research Question 5: By grade level within each cohort, what types of MAP and PSSA score-change patterns were exhibited by students?

Chapter 3

Methods

Data Source

The current study was conducted by collecting and analyzing shelf data from one suburban-rural, elementary school in northeastern Pennsylvania for seven cohorts of students who attended the third, fourth and fifth grades during the following five school years: 2005-2006, 2006-2007, 2007-2008, 2008-2009, and 2009-2010. Data were collected for those students with available Measures of Academic Progress (MAP) math domain scores, and Pennsylvania System of School Assessment (PSSA) math scores. The number of students enrolled in each grade in each school year and the number of students with usable data (students with at least a Fall or Spring MAP score and a spring PSSA score) are shown in Table 3.1. Students who had either no MAP scores and/or no PSSA score on file were excluded from the data base.

Table 3.1

Number of Students Enrolled and Number of Enrolled Students Included in Data Analyses

School year	Grade	Number of students enrolled	Number of students with usable data
2005-2006	3	130	104
2006-2007	3	140	122
2007-2008	3	123	111
2008-2009	3	102	93
2009-2010	3	122	112
2005-2006	4	132	122
2006-2007	4	132	124
2007-2008	4	132	126
2008-2009	4	99	97
2009-2010	4	104	95
2005-2006	5	132	125
2006-2007	5	124	119
2007-2008	5	137	132
2008-2009	5	117	108
2009-2010	5	101	100

Measures Included in the Analyses

The data used in this study consisted of scores from fall and spring administrations of the Math MAP Math domain and the spring administration of the PSSA Math.

Measures of Academic Progress

MAP is a computerized, standardized, adaptive assessment tool that dynamically measures students' performance by individually calibrating item selection for each student to determine a performance level (NWEA, 2003). If a student incorrectly answers a question, the subsequent question is slightly less challenging, or conversely, if a student correctly answers a question, the subsequent question is slightly more difficult. This process continues throughout the assessment, allowing for a specific measure of the student's actual achievement level. In addition to growth scale (Rasch unit) scores, the MAP results are interpreted as performance ranges (Below Basic, Basic, Proficient, and Advanced) that correspond to the performance level ranges of the PSSA.

The process of constructing the MAP involves several steps that include test design, definition of content, item selection, and test production (NWEA, 2003). MAP tests can be designed specifically for an agency, or school district, thereby allowing assessment of unique goals. Most MAP assessments include roughly four to eight goals, with five to six subgoals each that are typically based on state standards and are curriculum driven.

No time limit is set for completion of the MAP. Students are not permitted to skip any items and are unable to return to previously administered items. The assessment is designed for as many as four administrations per student per year. Upon completion of the test, the student's score and individualized goals appear on the screen. Reports can be generated for individual students, classes, grade levels, or entire districts. Scores are reported as Rasch unit scores, typically ranging between 150 and 300 points. Standard error of measurement is reported to be between 2.5 and 3.5 Rasch unit points. In addition to Rasch unit scores, percentile ranks are provided and collapsed into categories corresponding to PSSA performance categories.

The MAP demonstrates acceptable concurrent validity when compared to the PSSA for Grades 5 and 8, with a validity coefficient of .84 reported in technical descriptions of the MAPs psychometric characteristics (NWEA, 2003). Additionally, the Math MAP is reported as being highly and consistently correlated with other measures of academic achievement used by a variety of states. Studies regarding reliability for the MAP demonstrated strong findings with test-retest reliability in the spring of 2002, ranging from .84 to.91 for grades 2 throught10.

For the purpose of this study, the MAP is aligned with the Pennsylvania State Standards. The Assessment Anchor Content Standards (Assessment Anchors or Anchors) are organized into five content domains. These domains are similar to the five National Council of Teachers of Mathematics (NCTM) Standards and the five National Assessment of Educational Progress (NAEP) Reporting Categories. Pennsylvania Academic Standard Statements were examined and aligned with the NAEP Reporting Categories and NCTM Standards. At each grade level, the MAP and PSSA assess five skill domains of math: Numbers and Operations, Measurement, Data Analysis and Probability, and Algebraic Concepts (PA Standards 2.1-2.7, PDE, 2011). PA Standards

2.1 Numbers and 2.2 Computation are assessed by the category "Numbers and Operations", 2.3 Measurement is assessed by the category "Measurement", 2.9 Geometry and 2.10 Trigonometry are assessed by the category "Geometry;" 2.8 Algebra is assessed by the category "Algebraic Concepts", and 2.6 Statistics & Data and 2.7 Probability are assessed by the category "Data Analysis & Probability" (PDE, 2011).

Although it can be categorized as a CBM tool, the Math MAP assessment takes much longer to administer than do the other traditional CBM tools, such as AIMSweb. It is, however, both computer scored and administered and can be administered to students in groups, offering considerable advantages in terms of efficiency and convenience.

Pennsylvania System of School Assessment (PSSA)

The PSSA is a standards-based assessment that contains three content-specific assessments: reading, math, and writing. The current study focuses only on the math portion of the PSSA. It is administered in all public schools within the state of Pennsylvania for all students in Grades 3 through 8 and Grade 11 (PDE, 2007).

Student performance is reported in the form of scaled scores based on a mean of 1,300 and a standard deviation of 100. Scores also are categorized into one of four levels: Advanced, Proficient, Basic, and Below Basic. Advanced or Proficient performance indicates that a student has mastered Pennsylvania's assessment anchor content standards at their grade level (PDE, 2007).

The four performance levels are defined as follows (PDE, 2007):

Advanced. This level reflects superior academic performance. Advanced work indicates an in-depth understanding and exemplary display of the skills included in the Pennsylvania Academic Content Standards.

Proficient. Proficiency reflects satisfactory academic performance. Proficient work indicates a solid understanding and adequate display of the skills included in the Pennsylvania Academic Content Standards.

Basic. This level reflects marginal academic performance. Basic work indicates a partial understanding and limited display of the skills included in the Pennsylvania Academic Content Standards. This work is approaching satisfactory performance, but has not reached it. There is a need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient level.

Below Basic. The lowest level reflects inadequate academic performance. Below Basic work indicates little understanding and minimal display of the skills included in the Pennsylvania Academic Content Standards. There is a major need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient level.

Reliability coefficients of greater than 0.9 for PSSA math scores at all grade levels have been reported in technical publications (Thacker, 2004). High internal consistency estimates are believed to be the result in part, of the large number of test items included on the test (Thacker, 2004).

PSSA scores correlate positively and significantly with Scholastic Aptitude Test (SAT), Comprehensive Test of Basic Skills (CTBS)/Terra Nova, California Assessment Test, Version 5 (CAT-5), Northwest Evaluation Association (NWEA)tests, and New Standards Reference Exam (NSRE), with reported math score correlations typically ranging from 0.7 to 0.9 (Thacker, 2004).

Regarding socioeconomic status, research has indicated that economically disadvantaged students did not score as well as their peers on PSSA or comparison tests (Thacker, 2004).

Procedures

All MAP and PSSA scores were entered into separate grade and school year data files, and student names and school identification numbers were removed from the data files and replaced with study identification numbers to ensure confidentiality.

MAP Rasch unit scores and corresponding descriptive categories for fall and spring administrations were retained in the data files for statistical analyses. The four MAP category descriptors of Below Basic, Basic, Proficient, and Advanced were further collapsed into two categories: Not Proficient (a combination of the Below Basic and Basic categories) and Proficient (a combination of the Proficient and Advanced categories). PSSA scaled scores and descriptive category data were retained in the data files for statistical analyses. The four PSSA category descriptors of Below Basic, Basic, Proficient, and Advanced were further collapsed into two categories: Not Proficient (a combination of the Below Basic and Basic categories) and Proficient (a combination of the Proficient and Advanced categories).

Statistical Analyses

The relationship between Math MAP scores and descriptive categories and math PSSA scores and descriptive categories was examined using correlational and descriptive analysis techniques. Correlational analyses involved the calculation of correlations between MAP Rasch unit scale scores and PSSA scaled scores.

Descriptive analyses involved the following:

1. The construction of 2 x 2 cross-tabulation tables as shown in Figure 3.1 and the calculation of the following indices (also shown in Figure 3.1): Percentage of Students At-Risk, Percentage Change in Performance Category, Sensitivity, Specificity, Improvement, Instability and Kappa (representing predictive capacity beyond chance level).

2. The construction of status change patterns and categories as shown in Table 3.2.

Table 3.2

Construction of Cross-tabulation Tables and Indices Used in Statistical Analyses of Data

		PSSA Score Category	
		Not Proficient	Proficient
MAP Score Category	At-Risk or Not Proficient	A	B
	Not At-Risk or Proficient	C	D

Percentage of Students At-Risk = (A+C/(A+B+C+D)) x 100

Improvement Index = (B/(A+B)) x 100

Instability Index = (C/(C+D)) x 100

Sensitivity Index = (A/(A+C)) x 100

Specificity Index = (D/(B+D)) x 100

Kappa = ((po-pe)/(1-e)) x 100 where:

Po = pA +pD

Pe = ((pA +pC)(pA+pB)) + ((pB +pD)(pC+pD))

pA=A/Total *N* pB=B/Total *N* pC=c/Total *N* pD=D/Total *N*

A status change pattern was determined for each student by examining the descriptive categories obtained on the fall and spring administrations of the MAP and the spring administration of the PSSA and categorizing patterns of changes in status from Fall MAP to Spring MAP to Spring PSSA. Students were assigned to categories based on the pattern of relationship among these three scores as shown in Table 3.2. Percentages of students exhibiting each status change pattern were calculated for the score relationships at each grade level within each cohort.

Table 3.3

Status Change Patterns Within Status Change Categories

(Consistently Not Proficient Category)			
Pattern	Fall MAP	Spring MAP	PSSA
N-N-N	Not Proficient	Not Proficient	Not Proficient
Negative Change Pattern Category			
P – N – N	Proficient	Non-proficient	Non-proficient
P – P – N	Proficient	Proficient	Non-proficient
N – P – N	Non-proficient	Proficient	Non-proficient
Positive Change Pattern Category			
P – N – P	Proficient	Non-proficient	Proficient
N – N – P	Non-proficient	Non-proficient	Proficient
N – P – P	Non-proficient	Proficient	Proficient
Consistently Proficient Category			
P – P – P	Proficient	Proficient	Proficient

MAP and PSSA were analyzed according to student cohorts. Cohorts consisted of the grade level data sets that corresponded to a specific group of students as they moved from third to fourth to fifth grade. Because the data set included the test scores only from students in the third , fourth, and fifth grades in the specific school years of 2005-2006, 2006-2007, 2007-2008, 2008-2009, and 2009-2010, not all student cohorts included a complete set of files for third , fourth , and fifth grades. Table 3.3 details the number of grade years, school grade year(s), and school year(s) analyzed for each student cohort in this study. Two cohorts had only 1 year of data available for analysis. Cohort 1 had only 1 year of data available for analysis because the students of this cohort were in fifth grade during the 2005-2006 school year (the uppermost grade from the first year that was included in data used in the study). Cohort 7 had only 1 year of data available for analysis because the students of this cohort were in third grade in 2009-2010 (the final year of collected data that were used in the study). Cohorts 2 and 6 had only 2 years of data. Cohort 2 was in fourth grade in 2005-2006 and in fifth grade in 2006-2007. Cohort 6 was in third grade in 2008-2009 and in fourth grade in 2009-2010. Data analysis was possible for all three school grades, third, fourth, and fifth, for Cohorts 3, 4, and 5.

Table 3.4

Number of Grade Years Analyzed, School Grade Year and School Year Analyzed for Each Cohort

Year cohort entered kindergarten	Number of grade years analyzed for each cohort	School grade year(s) analyzed for each cohort	School year(s) analyzed for each cohort
2000-2001	1	Grade 5	2005-2006
2001-2002	2	Grades 4 and 5	2005-2006, 2006-2007
2002-2003	3	Grades 3, 4, and 5	2005-2006, 2006-2007, 2007-2008
2003-2004	3	Grades 3, 4, and 5	2006-2007, 2007-2008, 2008-2009
2004-2005	3	Grades 3, 4, and 5	2007-2008, 2008-2009, 2009-20010
2005-2006	2	Grades 3 and 4	2008-2009, 2009-2010
2006-2007	1	Grade 3	2009-2010

For purposes of data analysis reporting, most results tables are structured in a cohort-by-grade format for easier comparison of results across grades. This format allows for the most straightforward examination of same-grade data across cohorts and same-cohort data across grades. The format of these tables is shown in Table 3.4.

Table 3.5

Format for Reporting of Results by Cohort and by Grade

Cohort	Grade 3	Grade 4	Grade 5
1			X
2		X	X
3	X	X	X
4	X	X	X
5	X	X	X
6	X	X	
7	X		

Note. X = The cohort for the respective grade for that year.

Operational definitions for the indices and patterns used to analyze the data and interpret findings in this study are as follows:

Percentage of Students At Risk: The percent of students at risk is operationally defined as the percentage of students At-Risk of Not Being Proficient on the PSSA and MAP during that same school year.

Status Change Patterns and Categories: Status change patterns are based on the relationship among the score descriptive categories (Proficient or Not Proficient) assigned to a student's Fall MAP, Spring MAP, and PSSA performances. The eight possible status change patterns are shown in Table 3.2. The eight status change patterns are grouped into four status change categories: Consistently Not Proficient, Negative Change, Positive Change, and Consistently Proficient. The eight status change patterns and the four status change categories are defined as follows:

Consistently not proficient category and consistently not proficient pattern. For each cohort, the percentage of students who scored in the Not Proficient range on all three test administrations (Fall and Spring MAP, and PSSA*)* represents both the status change pattern and the status change category.

Negative change category. For each cohort, the percentages of students who exhibited negative change patterns of performance on the three assessments. The three negative change patterns identified for this study include the following:

P-N-N – the percentage of students who scored in the Proficient range on the fall administration of the MAP, then scored in the Not Proficient range on the spring administration of the MAP, and finally scored in the Not Proficient category on the PSSA.

P-P-N – the percentage of students who scored in the Proficient range on the fall and spring administration of the MAP but then scored in the Not Proficient range on the PSSA.

N-P-N – the percentage of students who scored in the Not Proficient range on the Fall MAP, then scored in the Proficient range on the Spring MAP, but then scored in the Not Proficient range on the PSSA.

Positive change. For each cohort, the percentages of students who exhibited positive change patterns of performance on the three assessments. The three positive change patterns identified for this study include:

P-N-P – the percentage of students who scored in the Not Proficient range on the Fall MAP, then scored in the Proficient range on the Spring MAP and PSSA.

N-N-P – the percentage of students who scored in the Not Proficient range on the fall and spring administration of the MAP but then scored in the Proficient range on the PSSA.

N-P-P – the percentage of students who scored in the Not Proficient range on the Fall MAP, then scored in the Proficient range on the Spring MAP and PSSA.

Improvement index. The Improvement Index is operationally defined as the percentage of students categorized as Not Proficient on the MAP but identified as Proficient on the PSSA. The Improvement Index represents the success rate of students identified as At-Risk of Being Not Proficient on the PSSA.

Instability index. The Instability Index is operationally defined as students who were identified as Proficient on the MAP who conversely earned scores in the Not Proficient range on the PSSA during that same school year.

Sensitivity. Sensitivity is operationally defined as the proportion of students who were identified as Not Proficient on the PSSA and also were identified as Not Proficient on the MAP during the same school year.

Specificity. Specificity is operationally defined as the proportion of students who were identified as Proficient on the PSSA and also were identified as Proficient on MAP during that same school year.

Kappa. The Kappa Index indicates the percentage of increase over chance level represented by the overall percentage of agreement of MAP and PSSA category level assignments during the same school year.

When calculating Improvement, Instability, Sensitivity, Specificity, and Kappa Index values for Fall MAP categories with PSSA categories and for Spring MAP categories with PSSA categories and when identifying status change patterns, analyses were conducted only with the data from students who had complete data sets within a grade level (i.e., only students who had taken all three tests – Fall MAP, Spring MAP, and PSSA during that school year). Although this inclusionary criterion eliminated a few students from each grade level data set for each cohort, it enabled meaningful comparisons of changes in index scores from fall to spring and meaningful interpretation of status change patterns within each grade level of each cohort. For this study, however, students within a cohort were not required to have complete data for each grade level.

This fact greatly constrains interpretation of indices and status change patterns across grade levels within a cohort, as each grade level analysis is composed of a somewhat different group of students at each grade level (i.e., students who were enrolled in that specific grade for that specific year and who took the Fall and/or Spring MAP and PSSA assessments that year). Although having complete data sets across each grade within each cohort is more desirable, keeping only students with complete data across all grades within a cohort would have resulted in eliminating from analyses a large percentage of the students in many of the cohorts, thereby severely constraining the interpretation of data at each grade level within each cohort. Table 3.5 shows the number of students included in data analyses for each grade level of each cohort and the number of students who would have been included in the data analyses of each cohort for each grade level had the analyses included only students enrolled at each grade level within each cohort.

Table 3.6

Number of Students Included in Grade Level Analyses and Number of Students That Would Have Been Included in Cross-Grade Level Analyses

	Grade Level *N* Counts of Students Enrolled in Each Grade			*N* Counts of Students Enrolled Consecutive Grades Within a Cohort		
Cohort	Grade 3	Grade 4	Grade 5	Grade 3-4	Grade 4-5	Grade 3-4-5
1	-	-	124	-	-	-
2	-	121	118	-	104	-
3	103	123	131	92	115	97
4	120	125	107	105	59	54
5	111	96	99	56	88	54
6	92	94	-	77	-	-
7	111	-	-	-	-	-

Chapter 4

Results

Research question 1. By grade level within each cohort, what is the relationship among scores on the fall and spring administrations of the MAP for math and the PSSA math assessment based on correlation coefficients?

Table 4.1 shows the results of correlational analyses comparing PSSA scores with Fall MAP and Spring MAP. Relatively high correlations were found between the Fall and Spring MAP and between the Spring MAP and PSSA. MAP correlations with PSSA are higher between the Spring MAP and PSSA than between the Fall MAP and PSSA. The lowest correlations are between Fall MAP and PSSA.

Table 4.1

Correlations Between MAP and PSSA Scores by Grade Within Cohort

	Grade 3			Grade 4			Grade 5		
Cohort	Fall MAP With Spring MAP	Fall MAP With PSSA	Spring MAP With PSSA	Fall MAP With Spring MAP	Fall MAP With PSSA	Spring MAP With PSSA	Fall MAP With Spring MAP	Fall MAP With PSSA	Spring MAP With PSSA
	Correlation Coefficient								
1							.85 (*n*=122)	.81 (*n*=122)	.80 (*n*=124)
2				.81 (*n* =114)	.81 (*n* =116)	.82 (*n* =119)	.84 (*n* =113)	.84 (*n* =114)	.84 (*n* =117)
3	.89 (*n* =97)	.80 (*n* =100)	.83 (*n* =100)	.88 (*n* =114)	.83 (*n* =115)	.86 (*n* =122)	.88 (*n* =127)	.87 (*n* =127)	.89 (*n* =127)
4	.61 (*n* =93)	.59 (*n* =116)	.82 (*n* =97)	.87 (*n* =122)	.85 (*n* =122)	.85 (*n* =125)	.87 (*n* =103)	.82 (*n* =104)	.85 (*n* =106)
5	.83 (*n* =99)	.80 (*n* =105)	.82 (*n* =105)	.87 (*n* =92)	.86 (*n* =94)	.88 (*n* =94)	.72 (*n* =76)	.63 (*n* =95)	.69 (*n* =80)
6	.78 (*n* =89)	.69 (*n* =90)	.75 (*n* =91)	.83 (*n* =82)	.54 (*n* =89)	.51 (*n* =87)			
7	.85 (*n* =103)	.67 (*n* =103)	.81 (*n* =111)						

Note: MAP = Measures of Academic Progress; PSSA = Pennsylvania System of School Assessment.

Research question 2. By grade level within each cohort, what proportion of students was identified as Proficient on the spring administration of the PSSA math assessment?

Table 4.2 shows the percentage of students who scored in the Proficient category on the Fall and Spring MAP and the PSSA. For each cohort and at each grade level, the percentage of students scoring in the Proficient range increased from the fall to the spring on the MAP. In addition, the percentage of students scoring in the Proficient range increased from the Spring MAP to the PSSA. The data demonstrate that student performance improved over the course of the year.

Table 4.2

Percentage of Students Earning Scores in the Proficient Range on the MAP and PSSA by Grade within Cohort

	Percent Proficient		
Cohort	Grade 3	Grade 4	Grade 5
1			72
			(*n* =124)
2		79	79
		(*n* =121)	(*n* =118)
3	87	81	71
	(*n* =103)	(*n* =123)	(*n* =131)
4	80	78	78
	(*n* =120)	(*n* =125)	(*n* =107)
5	88	83	80
	(*n* =111)	(*n* =97)	(*n* =99)
6	80	97	
	(*n* =92)	(*n* =94)	
7	86		
	(*n* =111)		

Note: MAP = Measures of Academic Progress; PSSA = Pennsylvania System of School Assessment.

Research question 3. By grade level within each cohort, what proportion of students was identified as Not Proficient on the Fall or Spring MAP math assessment and judged as At Risk of being Not Proficient on the PSSA math assessment?

Table. 4.3 shows the proportions of students who scored in the Not Proficient range on the Fall and Spring MAP, indicating that they were At Risk of being Not Proficient on the PSSA math assessment. A trend was seen across all cohorts and grade levels where greater proportions of students scored in the Not Proficient category on the Fall MAP than on the Spring MAP. The proportions of students who were found to be At Risk of being Not Proficient on the PSSA based on MAP scores therefore was less in the spring than in the fall.

Table 4.3

Percentage of Students Earning Scores in the Not Proficient Range (At Risk of Not Passing the PSSA) on the Fall and Spring MAP by Grade Within Cohort

Cohort	Grade 3		Grade 4		Grade 5	
	Fall MAP	Spring MAP	Fall MAP	Spring MAP	Fall MAP	Spring MAP
1					61	29
2			49	27	49	29
3	50	18	48	28	64	34
4	56	26	61	27	57	32
5	54	27	52	29	48	17
6	60	24	49	16		
7	53	19				

Research question 4. By grade level within each cohort, what is the relationship between the Fall and Spring MAP for math proficiency decisions and PSSA proficiency decisions based on multiple indices?

Cross-tabulation tables were constructed to analyze the relationship between Fall and Spring MAP and PSSA score levels in Grades 3, 4, and 5. The data from these cross-tabulation tables were utilized to calculate the Improvement, Instability, Sensitivity, Specificity, and Kappa Index values. Results of these analyses are reported in Tables 4.4 through 4.8 for each grade within the seven cohorts.

Research question 4a. At each grade within each cohort, of the students who earned Not Proficient level scores on the PSSA math assessment, what percentage of these students also earned Fall or Spring Math MAP scores in the Not Proficient range (operationally defined as the Sensitivity Index)?

Table 4.4 displays Sensitivity percentages from each cohort. These percentages indicate the percentage of students earning Not Proficient scores on the PSSA who also were categorized as At-Risk of Being Not Proficient or Not Proficient on the MAP. Separate percentages are reported for comparison of the PSSA with the Fall MAP and comparison of the PSSA with the Spring MAP. A consistent trend was observed across all grades of all cohorts whereby Sensitivity Index values were higher for PSSA compared with Fall MAP than for PSSA compared with Spring MAP. The proportion of student earning Not Proficient scores on the PSSA who also earned Not Proficient scores on the Fall MAP in relation to the total group of students who earned Not Proficient scores on the PSSA was greater than the number of students earning Not Proficient scores on the PSSA who also earned Not Proficient scores on the Spring MAP in relation to the total group of students who earned Not Proficient scores on the PSSA. Fall Sensitivity values ranged from 86 to 100, whereas Spring Sensitivity values ranged from 74 to 100.

Table 4.4

Sensitivity Index Values by Grade Within Cohort

Cohort	Grade 3		Grade 4		Grade 5	
	Fall	Spring	Fall	Spring	Fall	Spring
1					100	82
					(*n* = 122)	
2			100	83	91	74
			(*n* = 121)		(*n* = 118)	
3	100	91	100	91	86	78
	(*n* = 103)		(*n* = 123)		(*n* = 131)	
4	94	83	100	85	95	100
	(*n* = 93)		(*n* =125)		(*n* = 107)	
5	100	75	100	93	92	82
	(*n* = 111)		(*n* = 96)		(*n* = 99)	
6	94	83	100	100		
	(*n* = 92)		(*n* =94)			
7	100	85				
	(*n* = 111)					

Research question 4b. At each grade within each cohort, of the students who earned Proficient level scores on the PSSA math assessment, what percentage of these students also earned Fall or Spring Math MAP scores in the Proficient range (operationally defined as the Specificity Index)?

Table 4.5 displays Specificity percentages calculated for each grade within each cohort. These percentages indicate the percentage of students identified as Proficient on the PSSA who also were identified as Not At-Risk on the MAP. Separate percentages are reported for comparison of the PSSA with the Fall MAP and comparison of the PSSA with the Spring MAP. The Specificity Index reflects the level of agreement about positive outcomes on the progress-monitoring and PSSA measures. A consistent trend was observed for all grade levels within all cohorts whereby Specificity values were higher when comparing PSSA scores with Spring MAP scores than when comparing PSSA scores with Fall MAP scores. Spring Specificity values ranged from 82 to 93, whereas Fall Specificity values ranged from 49 to 66.

Table 4.5

Specificity Index Values by Grade Within Cohort

Cohort	Grade 3		Grade 4		Grade 5	
	Fall	Spring	Fall	Spring	Fall	Spring
1					53	91
					(*n* = 122)	
2			64	87	53	91
			(*n* = 121)		(*n* = 118)	
3	55	93	66	89	45	82
	(*n* = 103)		(*n* = 123)		(*n* = 131)	
4	57	88	49	90	53	88
	(*n* = 93)		(*n* = 125)		(*n* = 107)	
5	53	83	58	84	68	92
	(*n* = 111)		(*n* = 96)		(*n* = 99)	
6	49	92	55	86		
	(*n* = 92)		(*n* = 94)			
7	53	92				
	(*n* = 111)					

Research question 4c. At each grade within each cohort, what is the percentage of improvement over chance represented by the relationship between PSSA math assessment level score categories and Fall and Spring Math MAP score level categories (operationally defined as the Kappa Index)?

Table 4.6 displays Kappa Index values as percentages of agreement above chance between PSSA score levels and MAP score levels. Separate Kappa Index values are reported for comparison of the PSSA with the Fall MAP and comparison of the PSSA with the Spring MAP. The higher the Kappa Index value, the greater the agreement between PSSA and MAP score levels, regardless of whether score levels were Proficient or Not Proficient. A consistent trend was evident whereby Kappa Index values based on PSSA scores compared with Spring MAP scores were much higher than PSSA scores compared with Fall MAP scores. Spring Kappa Index values ranged from 23 to 75, whereas Fall Kappa Index values ranged from 6 to 44.

Table 4.6

Kappa Index Values by Grade Within Cohort

Cohort	Grade 3		Grade 4		Grade 5	
	Fall	Spring	Fall	Spring	Fall	Spring
1					39	72
					(*n* = 122)	
2			43	62	35	52
			(*n* = 121)		(*n* = 118)	
3	21	70	44	71	23	56
	(*n* = 103)		(*n* = 123)		(*n* = 131)	
4	31	63	29	68	30	75
	(*n* = 93)		(*n* = 125)		(*n* = 107)	
5	21	40	31	60	35	67
	(*n* = 111)		(*n* = 96)		(*n* = 99)	
6	25	70	6	23		
	(*n* = 92)		(*n* = 94)			
7	22	66				
	(*n* = 111)					

Research question 4d. At each grade within each cohort, of the students who earned Not Proficient level scores on the Fall and Spring MAP for math assessment, what percentage of these students also earned Spring PSSA math assessment scores in the Proficient range (operationally defined as the Improvement Index)?

Table 4.7 displays Improvement Index percentages based on comparing MAP and PSSA scores for each grade of each cohort. Separate Improvement Index values are reported for comparison of the Fall MAP with PSSA scores and comparison of the Spring MAP with PSSA scores. Improvement Index percentages indicate the percentage of students identified as At-Risk of Being Not Proficient on the PSSA based on Fall and Spring MAP score levels who performed at a Proficient level on the PSSA. A consistent trend was observed for all grades within all cohorts, whereby Improvement Index values were higher when comparing Fall MAP scores with PSSA scores than when comparing Spring MAP scores with PSSA scores. Improvement Index values ranged from 55 to 95 when comparing Fall MAP scores with PSSA scores, whereas values ranged from 22 to 85 when comparing Spring MAP scores with PSSA scores.

Table 4.7

Improvement Index Values by Grade Within Cohort

Cohort	Grade 3		Grade 4		Grade 5	
	Fall	Spring	Fall	Spring	Fall	Spring
1					55	22
					(*n* = 122)	
2			57	38	62	45
			(*n* = 121)		(*n* = 118)	
3	78	38	57	32	62	36
	(*n* = 103)		(*n* = 123)		(*n* = 131)	
4	65	38	65	31	64	31
	(*n* = 93)		(*n* = 125)		(*n* = 107)	
5	77	63	68	46	66	36
	(*n* = 111)		(*n* = 96)		(*n* = 99)	
6	68	29	95	85		
	(*n* = 92)		(*n* = 94)			
7	76	39				
	(*n* = 111)					

Research question 4e. At each grade within each cohort, of the students who earned Proficient level scores on the Fall and Spring MAP for math assessment, what percentage of these students also earned Spring PSSA math assessment scores in the Not Proficient range (operationally defined as the Instability Index)?

Table 4.8 displays Instability Index percentages based on comparing MAP and PSSA scores for each grade level within each cohort. Separate Improvement Index values are reported for comparison of the Fall MAP with PSSA scores and comparison of the Spring MAP with PSSA scores. Instability is defined as the percentage of students categorized as Not At-Risk on the MAP who earned PSSA scores in the Not Proficient range. A consistent trend was observed for most grades within all cohorts whereby Instability Index values were higher when comparing Spring MAP scores with PSSA scores than when comparing Fall MAP scores with PSSA scores. Instability Index values ranged from 0 to 11 when comparing Fall MAP scores with PSSA scores; values ranged from 0 to 10 when comparing Spring MAP scores with PSSA scores.

Table 4.8

Instability Index Values by Grade Within Cohort

Cohort	Grade 3		Grade 4		Grade 5	
	Fall	Spring	Fall	Spring	Fall	Spring
1					0	7
					(*n* = 122)	
2			0	5	3	7
			(*n* = 121)		(*n* = 118)	
3	0	1	0	2	11	10
	(*n* = 103)		(*n* = 123)		(*n* = 131)	
4	2	4	0	4	2	0
	(*n* = 93)		(*n* = 125)		(*n* = 107)	
5	0	4	0	2	2	3
	(*n* = 111)		(*n* = 96)		(*n* = 99)	
6	3	4	0	0		
	(*n* = 92)		(*n* = 94)			
7	0	2				
	(*n* = 111)					

Research question 5. By grade level within each cohort, what types of MAP and PSSA score-change patterns were exhibited by students?

Tables 4.9 through 4.15 display the percentage of students assigned to each score-change pattern within each cohort based on the pattern of proficiency status on the Fall and Spring MAP and the PSSA. The largest proportions were seen for students who consistently scored in the Proficient range on the Fall and Spring MAP and the PSSA. The patterns in the Positive Change category accounted for the second largest proportion of students for nearly all grade levels in all cohorts. Very low percentages of students scored in the Proficient range on the PSSA when they received Not Proficient ratings on both the Fall and Spring MAP. Low proportions of students changed categories from Not Proficient on both the Fall and Spring MAP to Proficient on the PSSA.

Table 4.9

Percent of Students in Each Change Category by Grade Within Cohort 1

	Grade 3	Grade 4	Grade 5 (*n* = 124)
Change Category	Percent in Each Category		
Consistently Not Proficient			
N – N – N			22.6
Negative Change			
P – N – N			0.0
P – P – N			0.0
N – P – N			4.8
Positive Change			
P – N – P			.8
N – N – P			5.6
N – P – P			27.4
Consistently Proficient			
P – P – P			37.1

Table 4.10

Percent of Students in Each Change Category by Grade Within the Cohort 2

	Grade 3	Grade 4 (*n* = 121)	Grade 5 (*n* = 118)
Change Category	Percent in Each Category		
Consistently Not Proficient			
N – N – N		16.5	13.6
Negative Change			
P – N – N		0.0	.8
P – P – N		0.0	.8
N – P – N		3.3	4.2
Positive Change			
P – N – P		1.7	2.5
N – N – P		8.3	9.3
N – P – P		18.2	19.5
Consistently Proficient			
P – P – P		46.3	44.9

Table 4.11

Percent of Students in Each Change Category by Grade Within the Cohort 3

Change Category	Grade 3 (*n* = 103)	Grade 4 (*n* = 123)	Grade 5 (*n* = 131)
	Percent in Each Category		
Consistently Not Proficient			
N – N – N	9.7	18.7	22.1
Negative Change			
P – N – N	0.0	0.0	0.0
P – P – N	0.0	0.0	3.8
N – P – N	1.0	1.6	2.3
Positive Change			
P – N – P	0.0	.8	.8
N – N – P	5.8	5.7	10.7
N – P – P	32.0	17.9	26.7
Consistently Proficient			
P – P – P	45.6	48.0	30.5

Table 4.12

Percent of Students in Each Change Category by Grade Within Cohort 4

	Grade 3	Grade 4	Grade 5
	(*n* = 93)	(*n* = 125)	(*n* = 107)
Change Category	Percent in Each Category		
Consistently Not Proficient			
N – N – N	16.1	17.6	22.3
Negative Change			
P – N – N	0.0	0.0	.9
P – P – N	1.4	0.0	0.0
N – P – N	2.1	3.2	0.0
Positive Change			
P – N – P	0.0	0.0	0.0
N – N – P	9.5	8.0	7.5
N – P – P	24.7	37.6	26.2
Consistently Proficient			
P – P – P	46.2	37.6	40.2

Table 4.13

Percent of Students in Each Change Category by Grade Within Cohort 5

	Grade 3	Grade 4	Grade 5
	(*n* = 111)	(*n* = 96)	(*n* = 99)
Change Category	Percent in Each Category		
Consistently Not Proficient			
N – N – N	8.1	15.6	9.1
Negative Change			
P – N – N	0.0	0.0	0.0
P – P – N	0.0	0.0	0.0
N – P – N	2.7	1.0	2.0
Positive Change			
P – N – P	.9	2.1	1.0
N – N – P	12.6	9.4	4.0
N – P – P	23.4	22.9	16.2
Consistently Proficient			
P – P – P	41.4	44.8	44.4

Table 4.14

Percent of Students in Each Change Category by Grade Within the Cohort 6

	Grade 3	Grade 4	Grade 5
	(*n* = 92)	(*n* = 94)	
Change Category	Percent in Each Category		
Consistently Not Proficient			
N – N – N	16.3	2.7	
Negative Change			
P – N – N	0.0	0.0	
P – P – N	0.0	0.0	
N – P – N	3.3	0.0	
Positive Change			
P – N – P	0.0	0.0	
N – N – P	6.5	11.7	
N – P – P	32.6	26.6	
Consistently Proficient			
P – P – P	38.0	46.8	

Table 4.15

Percent of Students in Each Change Category by Grade Within Cohort 7

	Grade 3	Grade 4	Grade 5
	(*n* = 111)		
Change Category	Percent in Each Category		
Consistently Not Proficient			
N – N – N	9.9		
Negative Change			
P – N – N	0.0		
P – P – N	0.0		
N – P – N	1.8		
Positive Change			
P – N – P	0.0		
N – N – P	6.3		
N – P – P	31.5		
Consistently Proficient			
P – P – P	43.2		

Chapter 5

Discussion

This study utilized data collected from a school district that administers the Math MAP as a diagnostic and progress-mon itoring tool. It was anticipated that this study would demonstrate that the MAP is a tool that can be used accurately and efficiently to identify students who are at risk for failure on high-stakes math achievement tests (PSSA math in this case). Efficient and accurate identification of students who are at risk for failure enables educators to provide academic interventions for those students. Educators can intervene with at-risk students with the goal of improving skills deficits, consequently improving performance on state-wide math tests.

Correlations Between MAP and PSSA Scores

Correlations among the Fall and Spring MAP and the PSSA produced relatively consistent results at most grade levels within all cohorts. In most cases, there was a high correlation (typically in the mid to high 80s) between Fall MAP and Spring MAP scores. Both Fall and Spring MAP scores tended to correlate moderately to highly with PSSA scores, with correlations ranging from the low 50s to mid 80s. MAP and PSSA scores obtained from more recent testing years tended not to correlate as highly as those obtained from earlier years. Spring MAP scores typically correlated slightly higher with PSSA scores than did Fall MAP scores. This result is anticipated if one assumes that academic intervention took place during the school year for those students who scored in the Not Proficient category on the Fall MAP. With academic intervention, one would expect to see at least a modest increase in student performance, thereby increasing the likelihood of a Spring MAP score that would be more in line with the spring administration of the PSSA. Based on this logic, however, one also would expect that the correlation between the Fall MAP scores and the Spring MAP scores would produce correlations similar in magnitude to those observed between the Fall MAP scores and the scores from the spring-administered PSSA. The consistently higher correlations between the Fall and Spring MAP scores compared to the correlations between the Fall MAP and PSSA scores likely are the result of the differences in the score metrics and the range of possible scores for the two tests. MAP scores typically ranged from 150 to 240, and score changes from fall to spring were relatively small in magnitude, even when much growth was observed. In contrast, PSSA scores typically ranged from 900 to 1,800, and score changes were relatively large in magnitude for many students.

Percentage of Students Scoring in the Proficient Range

Overall, the percentage of students who scored Proficient on the PSSA was relatively high for all grade levels within all cohorts. Fifth-grade passing rates tended to be lower than third-grade and fourth-grade passing rates. Third-grade passing rates ranged from 80% to 88%; fourth grade passing rates ranged from 79% to 97% and fifth-grade passing rates ranged from 71% to 80%. As noted in Chapter 3, definitive statements about changes in passing rates across grades within a specific cohort cannot be made because each grade level within a cohort was composed of a different group of students. Although year-to-year fluctuations were frequent, some specific trends were observed within grades across school years. After dropping from 79% to 71%, the fifth-grade data reflected consistent increases across the next 2years, improving from 71% to 78% and then from 78% to 80%. Similarly, after dropping from 81% passing to 78%, the fourth-grade data reflected increases across the next 2years, improving from 78% to 83% and from 83% to 97%. Although passing rates for third-grade students tended to be the highest in any given school year, the passing rates of third-grade students fluctuated up and down in each consecutive school year, producing no observable trend upward or downward.

Percentage of Students Earning Scores in the Not Proficient Range on the Fall and Spring MAP

Across all grade-level analyses, the proportion of students who scored in the Not Proficient category on the MAP decreased from the fall to the spring. MAP testing therefore, consistently indicated that fewer students were At Risk of Being Not Proficient (At Risk), on the PSSA in the spring than in the fall. Decreases in At Risk status were quite large at every grade level in every school year. The largest shifts in At Risk status were evident with third-grade classes, for which annual decreases in this status from Fall MAP testing to Spring MAP testing ranged from 29% to 36%. Although typically not as large as the decreases observed in third-grade classes, the annual decreases in MAP At Risk status ranged from 20% to 34%for fourth-grade classes and from 20% to 32% for fifth-grade classes.

One can hypothesize that curriculum changes and academic interventions that were implemented for those students who were identified as At Risk in the fall helped to improve their performance on the Spring MAP and subsequently on the PSSA. Although this hypothesis is reasonable, this study did not examine in detail the type of intervention provided to such students and did not utilize data sources sufficient to establish the efficacy of intervention efforts.

Comparing Progress Monitoring with PSSA Outcomes

Progress-monitoring results using the Math MAP in the fall and in the spring were compared to PSSA math results by comparing the proportions of students categorized as At Risk and Proficient on each measure and calculating values for five indices: the Sensitivity Index, the Specificity Index, the Kappa Index, the Improvement Index, and the Instability Index.

Sensitivity Index Results

The Sensitivity Index represents the percentage of students earning PSSA scores in the Not Proficient categories who also earned MAP progress-monitoring scores in the At-Risk categories. Ideally, all students who do not earn Proficient scores on the PSSA should be identified during progress monitoring as At-Risk; therefore, a Sensitivity Index value of 100% is the target for all educational programs. One must realize that Sensitivity Index values of 100% do not indicate anything about the number of students who are identified as At-Risk; rather, these numbers indicate the percentage of the students who did not pass the outcome measure (PSSA) who also were identified as At-Risk. If a program has only one student who fails the PSSA and that student had been identified as At-Risk, the Sensitivity Index value would be 100%. Likewise, the Sensitivity Index value would be 100% for a program in which 50% of the students do not earn Proficient ratings on the PSSA, and all of these students also had earned scores in the At-Risk range on the MAP. When progress is being monitored on a regular basis, Sensitivity Index values will fluctuate over time, consistent with fluctuations in category change values. Sensitivity Index values greater than 0 may indicate that students had not benefitted from general education or specific intervention efforts, or may indicate that students who had not benefitted from instruction were not able to demonstrate these skill gains on a standardized group assessment (PSSA).

Examination of Sensitivity Index in this study indicated a highly consistent pattern when Not Proficient status on the PSSA was predicted by Fall and Spring MAP. At all grade levels for nearly all years, Sensitivity Index values were higher when comparing PSSA outcomes to Fall MAP At-Risk status than when comparing PSSA outcomes to Spring MAP At-Risk status, although Sensitivity Index values remained very high even when comparing PSSA outcomes to Spring MAP results. The high Sensitivity Index values obtained when comparing PSSA outcome with Fall MAP results provide evidence that the MAP is a valid measure for identifying students at risk of failure on outcome measures such as the PSSA and that students who are identified as At Risk based on their Fall MAP performance are highly likely to be Not Proficient on their PSSA and, therefore, require academic intervention and math instruction. The fall to spring drops in Sensitivity Index values observed in most years for most grade levels indicate that in the spring, a greater number of students earning PSSA scores in the Not Proficient range earned scores in the Proficient or Not At Risk of Being Not Proficient (Not At-Risk) range on the Spring MAP, thereby reducing the overall accuracy of the MAP as a predictor of failure on the PSSA at that time of the year.

Specificity Index Results

The Specificity Index represents the percentage of students who earn PSSA scores in the Proficient category who also had earned MAP progress-monitoring scores in the Not At-Risk or Proficient categories. Ideally, similar to the Sensitivity Index, Specificity Index values would be 100%, thus indicating the percentage of students who were considered Not At-Risk or Proficient on the PSSA who also were considered Proficient on the MAP. All students who earn Proficient scores on the PSSA should have been identified during progress monitoring as likely to pass the PSSA (i.e., Not At-Risk). This Proficient score would demonstrate that these students maintained their Not At-Risk status, likely as a result of general education efforts or good test-taking skills. One must realize that Specificity Index values of 100% do not indicate anything about the number of students who are Not At-Risk or Proficient; rather, these numbers indicate the percentage of students who pass the outcome measure who also have performed effectively with the progress-monitoring measure. If 100 students pass the PSSA and all 100 had been predicted to pass the PSSA, the Specificity Index value would be 100%. However, the Specificity Index value would also be 100% for a program in which only 50 of the 100 students earn Proficient ratings on the PSSA, but all 50 of these students also had earned scores in the Not At-Risk range or Proficient range on a progress-monitoring measure. The Specificity Index answers the following question: "Of all the students who passed the PSSA, how many were predicted to pass?" The Specificity percentage, therefore, reflects only the level of agreement about positive outcomes between progress monitoring and PSSA measures. The Specificity Percentage does not provide information about the proportion of students who were identified as At-Risk on the progress-monitoring measure but who were able to earn scores in the Proficient range on the PSSA.

When MAP results were used in the current study as the indicator of At-Risk status, Specificity Index values were higher for the comparison with the Spring MAP than with the Fall MAP for all grade levels at every year. Specificity Index values ranged from only 45 to 68 for the Fall MAP compared to the PSSA but ranged from 82 to 93 for the Spring MAP compared to the PSSA. Ideally, the best pattern of results across multiple administrations of a progress-monitoring measure during the school year would be consistently high Specificity Index values, as close to 100% as possible. The pattern of lower Specificity Index values in the fall leading to much higher Specificity Index values in the spring, however, is not a surprising result. As the school year progressed, the accuracy of the prediction of students likely to perform in the Proficient range on the PSSA increased.

Kappa Index Results

The Kappa Index reflects the percentage of improvement over random assignment to Proficient and Not Proficient categories when progress-monitoring results are compared with PSSA results. The larger the Kappa Index value, the better the match between the progress-monitoring measure and the PSSA result. Comparisons of At-Risk status on progress monitoring measures with PSSA results should show less consistency in the fall than in the spring, when instructional efforts have had a longer time to impact student performance. Comparisons of At-Risk status on progress-monitoring measures with PSSA in the spring are less likely to show disagreement, as altering a student's status from At-Risk to Proficient is much more difficult to prior to PSSA testing because the time interval between progress monitoring and PSSA testing is much shorter. The accuracy of predictions (reflected in Kappa Index percentages) therefore should increase the closer the measure is administered to the PSSA.

Kappa Index values fluctuated for each grade level at each year. However, as predicted, the data show the trend that Kappa Index values are substantially greater for the comparison of the Spring MAP with the PSSA than for the Fall MAP with the PSSA. Kappa Index values were quite high when comparing Spring MAP with PSSA outcomes, generally ranging from 40 to 75. The only exception to this pattern was the lower value of 23 obtained for one fourth-grade class.

Ideally, progress-monitoring efforts should not demonstrate good long-term predictions (i.e. third-grade Fall MAP should not be a good predictor of fifth-grade PSSA performance) because that would mean that instructional efforts did not have any effect on the performance of students in the At-Risk category. The more predictive early measures are, the less productive education has been in meeting its purpose of increasing overall math proficiency. Conversely, short-term predictions should be more accurate because these measures would be given in closer proximity to PSSA administration, thus permitting less time for instruction or intervention to have a positive impact on At-Risk status.

Improvement Index Results

Improvement Index values indicate percentages of students identified as At-Risk by progress-monitoring measures but who are Proficient on the PSSA outcome measure.

Improvement Index values reflect a major change in status (from At-Risk to Proficient) and may be affected by general-education instruction or remedial math instruction.

At all grade levels at all years in the current study, Improvement Index values were higher when comparing Fall MAP with PSSA than when comparing Spring MAP with PSSA. The pattern of decreasing Improvement Index percentages is consistent with expectations from improvement efforts when the progress-monitoring measure is more closely aligned with the PSSA outcome measure. When alignment is high, the results obtained earlier in the year reflect not only a large gap between MAP results and PSSA outcome but also a long time during which intervention efforts can be applied to improve students' math skills. Conversely, the closer the MAP testing occurs to the time of the PSSA testing, such as in the spring, the more likely the PSSA outcome will mirror the MAP results, since the time for intervention efforts to effect a real change in math skill levels is much shorter, and Improvement Index values will be lower.

Instability Index Results

Instability Index values indicate the percentages of students identified as Not At-Risk or Proficient by a progress-monitoring measure who conversely earned scores in the Not Proficient range on the PSSA outcome measure. Ideally, no students identified as Not At-Risk or as Proficient would fail the PSSA outcome measure; an Instability Index value of 0%, therefore, is the target for all educational programs. This Instability Index value of 0% is important because a goal of education is to ensure that students do not fall from a Not At-Risk or Proficient status to a Not Proficient level on the PSSA.

Very low Instability Index values were observed at all grade levels across all years in the current study. For a majority of years at each grade level, Instability Index values were slightly higher when comparing PSSA outcomes to Spring MAP results than when comparing PSSA outcomes to Fall MAP results. In fact, the change in status from Proficient on the MAP to Not Proficient on the PSSA for such a small proportion of students as reflected by the very low Instability Index values could be attributed to the effects of random factors impacting on error of measurement, such as being sick or "having a bad day."

Score Change Patterns for Student Performance on the Fall and Spring MAP and the PSSA

Score status change patterns reflect each student's performance on all three assessments over the course of the year. Four status change categories were created to describe more concisely the various status change patterns observed in students' performances. The four status change pattern categories are Consistently Not Proficient, Negative Change, Positive Change, and Consistently Proficient. For every grade level in each cohort, the Consistently Proficient category always contained the largest percentage of students. The second largest percentage of students was always contained in the Positive Change category. Together, these two categories comprised between 63.7 % and 85.1% of each grade level's students, with values typically being in the 70% to 80% range. At each grade level in each year, Consistently Not Proficient students comprised between 2.7% and 22.6% of the students in the class, with typical values ranging from only 9% to 15%.

A very small proportion of students comprised the Negative Change category for each class, ranging from .9% to 6.8%, with most class percentages falling in the 1.0% to 3.0% range. The high percentages of students in the Consistently Proficient and Positive Change categories strongly suggests that educators were not only able to identify and intervene for students who were at risk for failure on the PSSA, but also able to ensure that a high percentage of students maintained positive growth that paid off with Proficient PSSA scores.

Summary

In the current study, relatively high correlations were found between the Fall and Spring MAP and the PSSA at many grade levels across many years. During some years for all grade levels, MAP correlations with PSSA were higher between the Spring MAP and PSSA than between the Fall MAP and PSSA. In other cases, only modest correlations were found between PSSA and both Fall and Spring MAP scores. The lowest correlations tended to be between Fall MAP and PSSA scores. Overall, the percentage of students who were Proficient on the PSSA was relatively high for all grade levels within all cohorts. Examination of Sensitivity Index indicated a highly consistent pattern when Not Proficient status on the PSSA was predicted by Fall and Spring MAP. At all grade levels for nearly all years, Sensitivity Index values were higher when comparing PSSA outcomes to Fall MAP At-Risk status than when comparing PSSA outcomes to Spring MAP At-Risk status, although Sensitivity Index values remained very high, even when comparing PSSA outcomes to Spring MAP results. Specificity Index values were higher for the comparison with the Spring MAP than with the Fall MAP for all grade levels at every year. The pattern of lower Specificity Index values in the fall leading to much higher Specificity Index values in the spring, however, is not a surprising result. The accuracy of the prediction of students likely to perform in the Proficient range on the PSSA increased when the MAP was administered closer to the time of administration of the PSSA. Kappa Index values fluctuated for each grade level at each year. However, as predicted, the data show the trend that Kappa Index values are substantially greater for the comparison of the Spring MAP with the PSSA than for the comparison of the Fall MAP with the PSSA. At all grade levels at all years, Improvement Index values were higher when comparing Fall MAP with PSSA than when comparing Spring MAP with PSSA. The pattern of decreasing Improvement Index percentages is consistent with expectations when the progress-monitoring measure is closely aligned with the outcome measure. When alignment is high, the results obtained earlier in the year reflect a larger gap between the progress-monitoring level and the outcome level than results obtained later in the year as there is much less time later in the year for intervention efforts to effect a real change in math skill levels.

Early intervention efforts are likely to improve the performance of students with fewer significant learning problems, leaving only the hardest-to-teach students in the progress-monitoring At-Risk category in the spring. Moving these hardest-to-teach students from At-Risk on the spring progress-monitoring to Proficient on the spring administration of the outcome measure is a much more difficult task than moving many mildly at-risk students into the Proficient range from the fall to the spring. Very low Instability Index values were observed at all grade levels across all years, indicating that very few students changed their status from Proficient on progress-monitoring measures to Not- Proficient on the outcome measure. This pattern of results is to be expected if general-education instruction is effective in providing students with the skills they need to maintain their Proficient levels of performance.

For a majority of years at each grade level, Instability Index values were slightly higher when comparing PSSA outcomes to Spring MAP results than when comparing PSSA outcomes to Fall MAP results. In fact, the change in status from Proficient on the MAP to Not Proficient on the PSSA for such a small proportion of students as reflected by the very low Instability Index values could be attributed to the effects of random factors impacting on error of measurement, such as being sick or “having a bad day.”

In terms of changes in performance across time within a school year, the Consistently Proficient category always contained the largest percentage of students for every grade level within every year. Likewise, the Positive Change category always contained the second largest proportion of students. A small proportion of students comprised the Consistently Not Proficient category for every grade level within every year, and a very small proportion of students exhibited Negative Change patterns for each class. The high percentages of students in the Consistently Proficient and Positive Change categories strongly suggest that educators were not only able to identify and intervene for students who were at risk for failure on the PSSA, but also able to ensure that a high percentage of students maintained positive growth that paid off with Proficient PSSA scores.

Limitations

A number of factors, including student population, district procedures, archival data sources, and data collection methods, can be viewed as limitations of the current study. Of greatest concern are limits placed on the generalizability of the study; generalization of results is limited based on the school district's demographics. Study results are applicable only to this specific school district and its student population. Specific demographic data for the student sample utilized in this study were not available and, therefore, not considered in the analyses of the data. Analyses were conducted only with the data from students who had complete data sets within a grade level (i.e., only students who had taken all three tests – Fall MAP, Spring MAP and PSSA during that school year). Although this inclusionary criterion eliminated a few students from each grade level data set for each cohort, it enabled meaningful comparisons of changes in index scores from fall to spring and meaningful interpretation of status change patterns within each grade level of each cohort.

For this study, however, students within a cohort were not required to have complete data for each grade level. Had such a matching procedure been implemented, as many as 20% to 65% of the students within cohorts would have been eliminated from the data analyses. This fact greatly constrained interpretation of indices and status change patterns across grade levels within cohorts, as each grade-level analysis is composed of a somewhat different group of students at each grade level.

An additional limitation was the fact that this study focused on only one Math CBM, the Math MAP. Additional types of Math CBMs warrant further investigation in order to determine which measure or measures are the strongest at predicting success or failure on state-wide achievement tests.

This study also did not measure response to instruction or interventions that may have been implemented by the school district nor how instruction or intervention impacted a student's performance on the PSSA. As a result, changes in student performance reflecting positive outcomes cannot be directly linked to instructional efforts of school staff. Although instructional efforts remain a likely source impacting on student performance, this study could not offer data to quantify or verify this impact.

Future Research

Relatively few studies to date have focused on Math CBMs and their ability to predict performance on state-wide achievement tests. Additional research is warranted with a variety of Math CBM tools in order to determine which Math CBMs are best at predicting success or failure on the PSSA and other state-wide math achievement tests. Future studies could focus on examining the effects of specific math general-education programs as well as remediation and intervention programs for those students found to be at risk of failure on state-wide math achievement tests. Identifying the most effective specific curriculum and remedial math programs for students found to be at risk for failure would prove very beneficial to students and school districts.

Implications for the Field of School Psychology

This study has added to the empirical research findings that CBM can be used as a predictor of state-wide assessment performance in math. The findings of this study support the contention that Math CBM measures can be used effectively as progress-monitoring measures and for predicting PSSA math performance.

Identifying students who are at risk for being not proficient on PSSA math allows for remediation of math skills and curriculum adaptations with the hopes of improving student performance on state-wide achievement tests. Early screening and progress-monitoring with Math CBM, across school years, may allow the use of appropriate interventions as early as possible. Improvements in screening and isolating students' math skill deficits could lead to improvement in their overall math achievement and increase the likelihood of earning scores of Proficient or "passing" on state-wide math achievement tests.

In addition to the standard methods for analyzing the relationship between screening measures and outcome measures that are typically used to assess the adequacy of progress-monitoring measures, this study utilized a number of unique data analysis techniques that were intended to increase the usefulness of analyses in determining the effectiveness of progress-monitoring measures. These techniques focused on showing the extent to which students improved, or did not improve, their performance on progress-monitoring measures across time and on determining how these changes related to performance on the outcome measure. These techniques included calculation of the Improvement and Instability Indices and calculation of status change patterns. The Improvement Index enables the researcher to identify the number of students At-Risk with a progress-monitoring measure who change their status to Proficient on an outcome measure. This proportion is important to take into account when evaluating the adequacy of instructional practices, as it indicates the number of students who were able to beat the prediction equation wherein poor performance on a progress-monitoring measure would predict poor performance on an outcome measure. When instructional and/or intervention efforts are carefully documented and students are monitored based on amount of instruction or intervention received, the Improvement Index can be used as an indicator of effectiveness of instruction, especially when a matched control design is implemented.

The Instability Index enables the researcher to identify the number of students as Not At-Risk with a progress-monitoring measure who change their status to Not Proficient on an outcome measure. This outcome is clearly undesired, as it reflects a negative change in student status. When Instability Index scores are very low, they are typically reflecting random fluctuations in students' test-taking performances. When Instability Index values increase into double digits, however, they are more likely reflecting a poor match between the content of the progress-monitoring measure and the outcome measure or, more importantly, the undesired effect of poor or inadequate instruction, inadequate curricula, and/or students' inability to profit from instruction, as a negative change in status not the result of fluctuation caused by sources other than a lack of content knowledge suggests a lack of adequate skill acquisition within the period between administration of the progress-monitoring measure and the outcome measure.

When Instability Index values are low, all students earning Not Proficient scores should be retested, as the most likely sources of the poor performance are factors not associated with a lack of content knowledge. When Instability Index values are high, the progress-monitoring measure and/or the math curriculum should be examined carefully to determine the extent to which they are aligned with the standards used to construct the outcome measure. This outcome measure is especially true if Improvement Index values also are very low.

The status change pattern analysis offers another more specific way to summarize data reflecting the pattern of change in performance across progress-monitoring measure and outcome measures. Dividing the status change patterns into four categories – Consistently Proficient, Positive Change, Negative Change, and Consistently Not Proficient – allows for a thorough analysis of progress-monitoring efforts and student performance patterns. Ideally, the percentage of students in the Consistently Not Proficient and Negative Change categories should be very small. The Consistently Proficient category reflects a high level of student knowledge and likely reflects a good match between instruction and intended outcomes and/or the presence of exceptionally talented students. The Positive Change category may reflect the results of good instructional efforts with students lacking in knowledge and/or positive changes in test-taking behavior attributable to factors other than instruction. The Negative Change category may reflect the results of poor instruction, a lack of match between progress monitoring and/or curricula with state standards used to develop the outcome measure, and/or negative changes in test-taking behavior attributable to factors other than poor instruction and/or standards mismatches. The Consistently Not Proficient category reflects a low level of student knowledge and may reflect a poor match between instruction and/or progress-monitoring measures and standards used to develop the outcome measure, poor instruction, and/or the presence of hard-to-teach students.

These new data analysis methods are intended to supplement the use of traditional methods, including the calculation of Sensitivity and Specificity Index values and the Kappa Index.

References

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L, Resetar, J. L., et al. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33*, 218-233.

Baenen, N., Ives, S., Lynn, A., Warren, T., Gilewicz, E., & Yaman, K. (2006). *Effective practices for at-risk elementary and middle school students* (E&R No. 06.03). Raleigh, NC: Wake County Public School System.

Baker, E. L., & Linn, R. L. (2003). Validity issues for accountability systems. In S. H. Fuhrman & R. F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 47–72). New York: Teachers College Press.

Barger, J. (2003). *Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina end of grade reading assessment* (Technical Report). Asheville, NC: North Carolina Teacher Academy.

Clarke, B., & Shinn M. R. (2004). A Preliminary Investigation Into the Identification and Development of Early Mathematics Curriculum-Based Measurement. *School Psychology Review 33*(2), 234-438.

Christ, T. J., & Vining, O. (2006). Curriculum based measurement procedures to develop multiple skill mathematics computation probes: Evaluation of random and stratified stimulus-set arrangements. *School Psychology Review, 35*, 387-400.

Clark, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum based measurement. *School Psychology Review, 33,* 234-248.

Cusumano, D. L. (2007). Is it Working? An Overview of Curriculum Based Measurement and its Uses for Assessing Instructional, Intervention, or Program Effectiveness. *Behavior Analyst Today, 8(1),* 24-34.

Deno, S. D. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Deno, S. D. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*, 184-192.

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19,* 1-16.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30(4)*, 507-524.

Deno, S. L. & Mirkin, P. K. (1977). *Data-based Program Modification: A Manual.* Reston VA: Council for Exceptional Children.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics. *The Journal of Special Education, 41*(2), 121-139.

ED Gov. (2006). *Building partnerships to help English language learners.* Retrieved from http://www.ed.gov/print/nclb/methods/english/lepfactsheet.html.

Elliot, S., & Fuchs, L. S. (1997). The utility of curriculum-based measurement and Performance assessment as alternatives to traditional achievement tests. *School Psychology Review, 26*, 224-234.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33,* 188-193.

Fuchs, L.S. & Deno, S.L. (1992). Effects of curriculum-based measurement. *Exceptional Children, 58*, 232-243.

Fuchs, D., & Fuchs, L.S. (2001). Helping Teacher Formulate Sound Test Accommodations Decisions for Student with Learning Disabilities. *Learning Disabilities Research & Practice, 16*(3), 174-181.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93-99. doi:10.1598/RRQ.41.1.4

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlet, C. L., & Seethhaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children, 73,* 311-330.

Fuchs, L. S., Fuchs, D., & Courey, S. J. (2005). Curriculum-based measurements of mathematics competence: From computation to concepts and applications to real-life problem solving. *Assessment for Effective Instruction, 30,* 33-46.

Fuchs, L.S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Classwide curriculum-based measurement: helping general educators meet the challenge of student diversity. *Exceptional Children, 60*(6), 518-537.

Graves, A. W., Plasencia-Peinado, J., Deno, S. L., & Johnson, J. R. (2005). Formatively evaluating the reading progress of first-grade English learners in multiple-language classrooms. *Remedial and Special Education, 26,* 215-225.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review, 34,* 372-386.

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review, 34,* 9-26.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33,* 193-203.

Messick, S. (1984). Assessment in context: Appraising student performance in relation to instructional quality. *Educational Researcher, 13*, 3-8.

Morison, P., White, S. H., & Feuer, M. J. (Eds.). (1996). *The use of IQ tests in special education decision making and planning: Summary of two workshops.* Washington, DC: National Academy Press.

National Center for Education Statistics. (2003). *Highlights From the Trends in International Mathematics and Science Study (TIMSS).* Retrieved from www.nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2013009.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, H.R. 1, 115 Stat. 1425 (2002, January 8). Retrieved from www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf.

Northwest Evaluation Association. (2003). *Technical manual for the Northwest Evaluation Association (NWEA) Measures of Academic Progress and Achievement Level Tests.* Portland, OR: Northwest Evaluation Association.

Pennsylvania Department of Education. (2002). *Handbook for report interpretation: 2002 PSSA mathematics and reading assessment for grades 5, 8, and 11*. Harrisburg, PA: Author.

Pennsylvania Department of Education. (2007). *2006-2007 PSSA results frequently asked questions.* Retrieved from http://www.pde.state.pa.us

Pennsylvania Department of Education. (2011). *Pennsylvania System of School Assessment:2010-2011 Assessment Handbook*. Harrisburg, PA: Retrieved from www.education.state.pa.us

Pennsylvania Department of Education. (2003, February). *Technical analysis: Pennsylvania system of school assessment, 2002 reading and mathematics PSSA.* Harrisburg, PA: Author.

Prewitt, K. (2002, Winter). *Demography, diversity and democracy.* The 2000 Census story. In K. Prewitt (Ed.), Brookings Review (pp. 6–9). Washington, DC: Brookings Institute.

Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). NAEP *1996 Mathematics Report Card for the Nation and the States.* Washington, DC: National Center for Education Statistics.

Sattler, J. M. (2001). Assessment *of children: Cognitive applications, Fourth edition.* San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L.E., & Hintze, J. M. (2006).Curriculum based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*, 19-35. doi:10.1177/0734282905285237

Shaw, R. & Shaw, D. (2002). DIBELS *Oral Reading Fluency-Based Indicators of Third Grade Reading Skills for Colorado State Assessment Program (CSAP)* (Technical Report). Eugene, OR: University of Oregon.

Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children*. New York; Guilford.

Speece, D. L., Case, L. P., & Molloy, D. W. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research & Practice, 18,* 147–156.

Stage, S., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30,* 407-419.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42(8)*, 795-819.

Thacker, A. A. (2004). *PSSA issues and recommendations (Human Resources Research Organization Report No. FR-04-34).* Alexandria, VA.

U.S. Department of Education, National Center for Education Statistics. (1998). *Children's Reading and Mathematics Achievement in Kindergarten and First Grade*, NCES 2002-125, Washington, DC.

U. S. Department of Labor, Bureau of Labor Statistics. (1997). *Occupational outlook handbook.* Washington DC: Government Printing Office.

Wallace, T., Espin, C. A., McMaster, K., Deno, S. L., & Foegen, A. (2007). CBM progress monitoring within a standards based system. *The Journal of Special Education, 41*(2), 66-67.

Wood, D. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11,* 85-104.